

HOW OLD ARE RNA NETWORKS?

Toni Daly,^{1,3} X. Sylvia Chen² and David Penny^{*,3}

¹Allan Wilson Centre of Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand;

²Department of Biochemistry, University of Otago, Dunedin, New Zealand; ³Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

*Corresponding Author: David Penny—Email: d.penny@massey.ac.nz

Abstract: Some major classes of RNAs (such as mRNA, rRNA, tRNA and RNase P) are ubiquitous in all living systems so are inferred to have arisen early during the origin of life. However, the situation is not so clear for the system of RNA regulatory networks that continue to be uncovered, especially in eukaryotes. It is increasingly being recognised that networks of small RNAs are important for regulation in all cells, but it is not certain whether the origin of these networks are as old as rRNAs and tRNA. Another group of ncRNAs, including snoRNAs, occurs mainly in archaea and eukaryotes and their ultimate origin is less certain, although perhaps the simplest hypothesis is that they were present in earlier stages of life and were lost from bacteria. Some RNA networks may trace back to an early stage when there was just RNA and proteins, the RNP-world; before DNA.

INTRODUCTION

With new classes of RNA continuing to be discovered, it appears as if many classes of RNA are likely to occur in eukaryotes, bacteria and archaea—but what about the actual RNA-protein networks in which they are involved. For example, small RNA regulation of gene expression has been seen in both eukaryotes and prokaryotes and even viruses use small RNAs. Although the basic mechanism of target recognition and cleavage is similar in all these groups, the proteins and interactions differ. This chapter considers some ideas for the time of origin of some key classes of RNA and their associated networks. We readily accept the idea that some RNA-protein interactions are very old but when we get down to the elaborate pathways of RNA-based regulation there was an early assumption that such regulation only evolved when organisms became more complex (e.g., multicellular).

Now we see that the networks of RNA-protein interaction are more general; but can we infer the presence of particular RNA functions in **Ida** (Initial Darwinian Ancestor), or in the later **Luca** (the Last Universal Common Ancestor), or in the even later **Fred** (Fairly Remote Eukaryotic Daddy).^{1,2}

When it comes to understanding the main roles of RNA in modern cell biology there is a potential problem of the ‘alphabet soup’ formed from so many classes of small RNAs. Defining these subgroups is important for identifying subclasses that are reasonably closely related—they have well-defined homologies and clearly related functions as collated and updated in the RNA database Rfam.³ However, our interest here is at the other end of the spectrum—once we have these many classes and networks of small RNAs, can we put them into higher level evolutionary groups with more general functions. As an example, Boria et al⁴ identified sbRNAs (stem bulge RNAs) in nematodes as evolutionary homologues of the more widespread Y-RNAs. Y-RNAs are involved with the protein R_0 in a network of interactions that both assist misfolded RNAs but are also involved with DNA transcription. This equivalence of R_0 and sbRNAs reinforces the message that RNA networks evolve, and as we understand more about how RNA and its associated proteins evolve, we can begin to surmise how such regulation could have evolved much earlier in the beginning of life.

REGULATORY NETWORKS OF SMALL RNAs

Networks involving small RNAs can regulate translation with some proteins also used in defence networks. Small (~20-30 nt) RNAs have many functions including, fine-tuning expression of temporal and tissue specific mRNA, destruction of aberrant mRNAs via cleavage, repression, up-regulation and also translational control via DNA methylation and maintenance of histone architecture.

Three basic regulatory classes of ncRNA networks are recognized in eukaryotes, determined by their biogenesis and mode of operation⁵ (siRNA-based, miRNA-based and piRNA-based). miRNA sequences are usually found in intergenic and intronic regions of eukaryotes and in some viruses. They have their own promoters and regulatory units. RNA Pol II transcribes long unstable primary RNAs that form hairpin loops (pri-miRNA). Some miRNA genes are found in the UTR regions of coding sequences (indicating that the transcript can be processed as either a miRNA or an mRNA) and some are even found in exons. A complex cellular milieu of miRNAs allows tuning of thousands of genes through combinatorial interactions within 3' untranslated regions and could account for cell specificity and/or temporal control.⁶ miRNAs were once thought to be cell-specific but recent work has shown that in plants at least, miRNAs move extensively between cells and can control protein levels in remote tissue.⁷ In a viral example, the herpes virus (dsDNA) produces a range of pri-miRNA transcripts which are processed to pre-miRNAs in the host nucleus. They utilise the endogenous miRNA network using the host RNase III endonuclease (Drosha) and a dsRNA binding protein (Pasha). These are then exported via the same pathway as endogenous miRNAs. These viral miRNAs can extend the lytic phase of the cell, by suppression of apoptosis. Other DNA viruses seem to produce just one or two miRNAs.⁸ miRNAs have an obligate nuclear processing phase, so viruses that enter the nucleus are more likely to encode them. HIV is an RNA retrovirus that encodes at least three miRNAs.^{9,10} Similarly with bacteria, the success of

the plant bacterium *Agrobacterium tumefaciens* is enhanced by knocking out the siRNA pathway, but still requires an intact plant miRNA pathway.¹¹

In animals, the network of miRNA-based reactions involves exportin-5 mediating nuclear export of pre-miRNAs, a cytoplasmic RNase III endonuclease Dicer, together with the dsRBP Loqs, (*Drosophila*) and trans-activator binding protein (TRBP) (in humans), transforms the pre-miRNA into the mature transcript. Plants lack a Drosha and Pasha homologue and instead use a nuclear Dicer-like endonuclease (DCL1), which makes similar cuts, but they also require a dsRBP (HYL1) to accurately process the miRNA precursors.¹² Such changes over time are what we expect from an evolutionary process and this makes it likely that the plant/animal ancestor had a comparable system.

Animal miRNAs often have limited complementarity, either as a duplex (miRNA:miRNA*) or with the target mRNA. Usually the strand with the lower 5' end stability enters the Argonaute domain of the RISC complex (RNA induced silencing complex). Argonautes are an ancient protein family known as 'slicers'. Plant and animal miRNAs usually load into Ago1 although miRNAs have been found in all four human Agos. It was believed that the miRNA* strand is degraded, but recent work in *Drosophila* indicates that the miRNA* strand can be modified by 2'-O-methylation at the 3' end then loaded into Ago2 in the RISC complex, the domain usually occupied by siRNA.¹³

In animals, introns can be linearized by the lariat debranching enzyme and the resulting RNA folds to a pre-miRNA structure. These spliced lariats are processed by Dicer and have been termed mirtrons. They have the same action as a miRNA, but do not require the cleavage capabilities of Drosha. Initially discovered in *D. melanogaster* and *C. elegans*, mirtrons have now been found in mammals, including primate specific sets.¹⁴ Additionally the HIV virus encodes a miRNA; hiv1-miR-TAR that uses the cellular mirtron pathway.⁹ It has been suggested¹⁴ that this pathway could predate Drosha mediated cleavage. An alternate would be that they have evolved independently three times in different lineages and that viruses have taken advantage of both pathways. Most animal Agos now lack cleavage ability¹⁵ and suppress mRNA translation by physical impedence, or by suppression of transcription via heterochromatin and DNA methylation.¹⁶

Animal miRNAs usually suppress translation by binding to the target mRNA, but the exact mechanism remains under debate. Three models¹⁷ for preventing initiation are; miRNA:RISC complexes compete for ribosome binding, compete for cap binding, or prevent circularization of the mRNA by inducing de-adenylation (circularization enhances mRNA translation). The latter two strategies could also promote the degradation of mRNA unprotected by cap or polyA tail. Another suggestion is that even if translation has begun, the miRNA:RISC complex could cause ribosomes to drop off prematurely. However, miRNAs can also up-regulate translation (particularly when the cell is under stress).¹⁸ Binding position is relevant—an example being if the interaction of mammalian miR10a:RISC is within the 3'UTR of ribosomal proteins then miR10a suppresses translation, but in the 5'UTR it could activate it.¹⁹ Thus there is a wide diversity of miRNA processing and the existence of small RNA networks appears universal in eukaryotes.

Recently small RNAs derived from tRNA were discovered.²⁰ These load preferentially into Ago3 and Ago4 where one of the classes of tRNA derived small RNAs (tsRNA) has a moderate capability to down-regulate mRNA. Ago3 and Ago4 may act as a buffer by loading unstructured small RNAs preventing Ago2 from becoming overloaded.²¹ Intriguingly they are confined to the cytoplasm, though their processing signature indicates that they are processed in the nucleus. Cellular localisation also plays a part in the standard miRNA-processing network. The miRNA:RISC complex in the cytoplasm

often relocates to P bodies enriched in mRNA degradation proteins.²² Smaller complexes composed of just siRNA or miRNA and Ago2 can also be found in the nucleus. Possibly, a small proportion of the cytoplasmic Ago2:RISC complex is stripped of Dicer and TRBP prior to, or during translocation to the nucleus. In the case of siRNA or perfectly complementary miRNAs, the target is cleaved and the Ago2 is released and exported to the cytoplasm. In *C. elegans*,²³ NRDE3 (an Ago) together with siRNA is necessary and sufficient for location to the nucleus and silencing of nuclear mRNAs. Although there are 27 Agos in nematodes, there appear to be little redundancy in that mutants lacking this protein (*nrde3*- nematodes) are defective for all nuclear mRNA silencing.

miRNA-mediated cleavage appears similar to the RNA interference mechanism mediated by siRNAs (see later). However, there are subtle differences in their requirement for complementarity; siRNA and plant miRNAs have perfect complementarity to their target, but the animal miRNA:miRNA* duplex contains mismatches bulges and GU wobble pairs. Their biogenesis is more strikingly different; most siRNAs silence their encoded targets and are processed from complementary dsRNA transcripts from invasive nucleic acid during infection (exo-siRNA) and so processing is often in the cytoplasm. In contrast, miRNA sequences are encoded in the organism's genome and so the first steps in the processing pathway occur in the nucleus and requires export. In plants and worms, each siRNA precursor gives rise to many siRNA duplexes, but only one miRNA:miRNA* is generated from each pre-miRNA. Additionally miRNAs are nearly always conserved in related organisms but exo-siRNAs are rarely conserved.²⁴

Endogenous-siRNA (endo-siRNA) is involved in a pathway that could evolutionarily link the nuclear and cytoplasmic processing. This process offers protection from genomically integrated parasitic DNA so its transcripts are derived from the genome and processed in the nucleus. The aim is the suppression of parasitic DNA either by guiding histone modification and DNA methylation, or cleavage of mRNA. The endo-siRNA pathway shares some of the miRNA proteins. For instance, *Drosophila* uses the double stranded RNA binding protein (dsRBP) Loquacious (Loqs) and Dicer 2 for the endo-siRNA pathway and another isoform of Loqs is used for the miRNA pathway in partnership with Dicer 1.²⁵ The commonality is the endogenous root of the RNA, but the need to quell exogenous transcripts would likely have arisen before silencing one's own.

Piwi interacting RNAs (piRNAs) were originally detected in *Drosophila* germline cells and they are slightly larger (~24-31 nt) RNAs with the 2' O-methyl 3' ending reminiscent of siRNAs. They are particularly known for the suppression of transposons and repeat sequences (see later), but flies and vertebrates have an additional type of piRNA expressed in germline cells during the pachytene and prepachytene stages of meiosis (known as primary piRNA). They are derived from the 3' UTR of cellular transcripts involved in diverse cellular processes. They do not exhibit the 'ping-pong' type amplification where secondary piRNAs are produced via an Aub/Ago3 dependent loop as seen in the suppression of transposons (see later), and they are depleted for repeat elements. The mechanism of primary piRNA production is poorly understood but it is thought that they are generated at the same time as the mRNA is being translated in the cytoplasm²⁶ and may be derived directly from mRNAs selected for piRNA production. There is some crossover of pathways here because in *Drosophila* at least, primary piRNAs that are involved with the suppression of transposable elements have been found in the somatic cells surrounding germ cells. These cells lack Aub and Ago3 and appear to come from one cluster (flamenco) and target one class of transposon (gypsy elements). Gypsy elements mobilise initially by being copied into RNA, but are then processed into virus-like particles

in the cytoplasm.^{27,28} It is intriguing that the siRNA, pathway which specialises in dealing with viral RNA, appears to have neglected gypsy elements and that in flies at least, a variant of the piRNA pathway has developed to deal with them. The important conclusion is again that there are many variations in details within eukaryotes, but the same basic mechanism seems to re-occur.

RNA REGULATION AND DEFENCE AGAINST THE DARK ARTS

All cells and even viruses, are subject to de novo invasion by both exogenous and endogenous nucleic acids including transposons, viruses, pseudogenes—essentially what has been called ‘selfish DNA’. We refer to this potential invasion of DNA as the ‘Dark Arts’ and (with acknowledgement to JK Rowling) ‘Defence against the Dark Arts’ heavily involves RNA networks.²⁹ Exogenous nucleic acid can occasionally be advantageous for an organism, but in most cases it will be deleterious and possibly lethal.

We first discuss the small RNAs involved in RNA interference (RNAi) pathways in eukaryotes and viruses and the following section the clustered, regularly interspaced, short palindromic repeats (CRISPR arrays) found in bacteria and archaea. We expect that at all stages during the origin of life, the same basic biological principles apply,³⁰ and that parasites would always have been present.³¹ Viruses have also evolved small-RNA based mechanisms which use the conserved RNAi machinery of the host.⁸ The co-evolution of these mechanisms is important for understanding molecular epidemiology ‘in action’.

We consider two forms of siRNA networks as described earlier; exogenous siRNAs (exo-siRNA) from invasive transcripts and short RNAs produced from endogenous transcripts (endo-siRNA). Exo-siRNAs are processed from long double stranded RNAs (dsRNA), often transcribed from viral or plasmid sources during infection in animals, fungi, protists and plants. It forms a basis for antiviral defence and requires at least three key proteins, the RNaseIII endonuclease—Dicer, an RNA dependant RNA polymerase (RdRP) and a member of the Argonaute (Ago) family.

There are 3 clades of Argonaute; ‘Argonaute-like’, ‘Piwi-like’ (both found in prokaryotes) and ‘group 3 Argonautes’ found in *C. elegans*.³² The Ago proteins are found in all three domains, but are quite diverse in sequence. They possibly originated as a structural support for catalytic RNA and eventually took on the catalytic role leaving the RNA as a guide? There is large variation between eukaryotes in their complement of these proteins (see Table 1) but our main question here is whether there are basic mechanisms across all eukaryotes. To be more precise, did the last common ancestor of eukaryotes (Fred) have this defence mechanism, which has subsequently been modified in the different eukaryote lineages? It has been unclear whether mammals use the exo-siR system given the complex immune system available to them. There is some evidence that mammalian viruses encode RNAi suppressors and that implies that they could use exo-siR,³³ and more recently virally encoded siRNAs have been found in mammalian cells. For example, the abundance of virally encoded siRNAs increases in cells defective for the interferon pathway (IFN α/β) (initially thought to preclude the possibility of exo-siRNA in mammals).³⁴

Exo-siRNA deals with invading RNA, but eukaryote genomes are largely comprised of selfish DNA that has accumulated over time and a variant of the siRNA system (endo-siRNA) deals with this. In plants RDR2 copies transcripts from silent loci, principally transposons, thought to be produced by RNA pol IV. This produces long dsRNA which is

Table 1. Some of the key enzymes in the RNAi pathways and variation between species

Species	RNase III Endonuclease	RNA Dependent RNA Polymerase	Argonaute-Like	Piwi-Like
<i>S. pombe</i>	DCR1	RDRP1	Ago1	
<i>A. thaliensis</i>	DCL1-4	RDR1, RDR2, RDR6	Ago1-10	
<i>D. melongaster</i>	DCR1-2, Drosha	A subunit of RNA pol II has RdRP capability	Ago1-2	Piwi, Aub, Ago3
<i>C. elegans</i>	DCR1, Drosha	Rrf-1, Rrf-3	Alg1-2, T22B3.2, T23D8.7, ZK757.3	PRG1-2, ERGO1 Plus at least 18 group 3 Argonautes
<i>H. sapiens</i>	DCR1, Drosha	RdRP composed of hTERT in complex with <i>RMRP</i> ³⁷	Ago1-4	HILI, HIWI, HIWI2 PIWIL3

cleaved to make sets of cis-acting siRNA (casiRNA) that affect the transcripts of the gene loci that produced them. RDR6 can copy miRNAs from the *TAS* locus, which then enters the RNAi pathway and is trans-acting (tasiRNA); i.e., produced from a discrete locus. Plants also have a third method of producing siRNA via RDR6 by copying overlapping transcripts, one produced constitutively and one produced in times of biotic stress. These are natural antisense transcript-derived siRNAs (natsiRNA).³⁵ This continued evolution of RNA-protein networks makes it harder to recognise specific networks, but does not obscure the general mechanisms of defensive RNA networks.

Classical RdRP homologs have not been found in mammals, yet endo-siRNAs appear ubiquitous in eukaryotes. This defence is important in flies, and an elongation subunit of RNA pol II has been found to have RdRP capability and is involved in the RNAi pathway.³⁶ An alternate RdRP is also available to mammals; an RdRP composed of hTERT in complex with *RMRP* has been shown to produce siRNAs in humans.³⁷ Endogenous dsRNA can also derive from DNA transcripts that would result in RNA prone to forming duplexes (Fig. 1A),³⁸ e.g., bi-directional transcription of a single gene, trans-interaction between transcripts from a founder gene and pseudogene, or overlapping transcription across two genes (convergent transcription). Long single stranded RNA transcripts that fold back on themselves could form from long transcripts of transposons in inverted orientation, or from duplicated inverted pseudogenes, or from structured loci. All of these could result in hairpin duplex structures that can be processed using the siRNA pathway with the addition of a double stranded RNA binding protein (dsRBP) (Fig. 1B).

In relation to the Dark Arts, exo-siRNA defends against exogenous viral and plasmid nucleic acid infection and endo-siRNAs against RNA from endogenous repeating units, transposons, integrated viral sequences and pseudogenes. This defence can occur in any cell, but is especially important in organisms (such as animals) where there is a division into somatic and germ-line cells. The matched endogenous repeat sequences were originally termed repeat associated small interfering RNAs (rasiRNA). This term has been retained for plant endo-siRNAs that target repeat sequences. However, piRNAs aren't simply longer siRNAs—their biogenesis is different, because they are independent of Drosha and Dicer endonucleases and they interact with a specific clade of Ago proteins

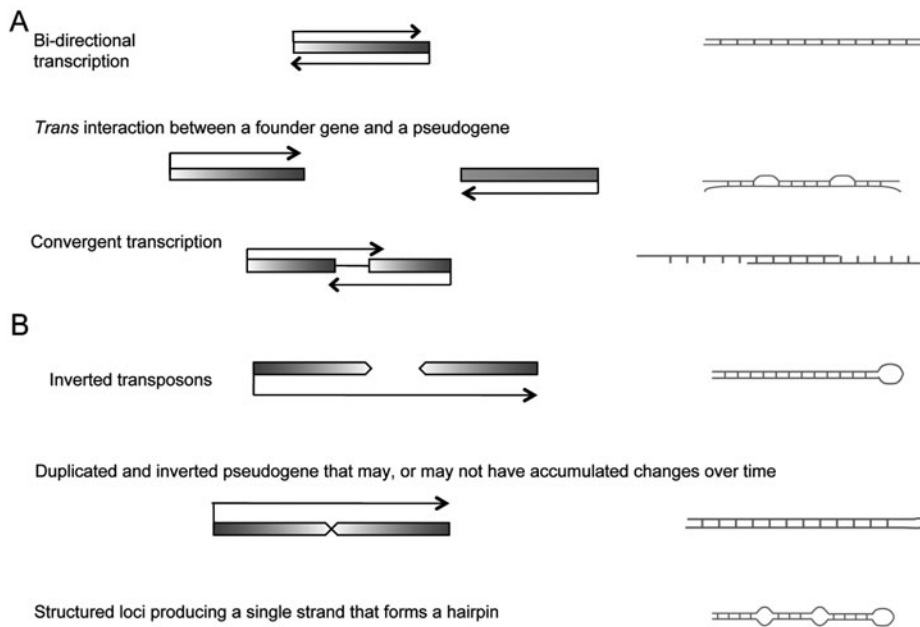


Figure 1. A) Transcription of endogenous DNA that gives rise to dsRNA. Bidirectional transcription resulting in dsRNA can arise when sense and antisense promoters are found in the 5'UTR. Parent and pseudogene transcripts can form endo-siRNAs that regulate the parent gene. Pseudogenes are essentially unsuccessful duplications, 'faulty' in some way, yet many are retained and don't appear to accumulate as many mutations as would be expected. siRNAs can also derive from overlapping transcripts, this occurs naturally in plants giving rise to natsiRNAs produced at times of stress. Convergent transcription in the geminivirus gives rise to a very short overlap but is sufficient to produce the dsRNA needed for exo-siRNA. B) Transcripts prone to forming hairpin loops form from the 'read through' of transposons or pseudogenes that have duplicated and inverted. A color version of this figure is available at www.landesbioscience.com/curie.

known as Piwi, AUB (Aubergine) and Ago3 in *Drosophila* (MILI1, MIWI and MIWI2 in mice, HILI, HIWI1, HIWI2 and HIWI3 in humans). The evolution of RNA networks continues; it is not a 'once and for all time' setup.

Transposons make up the bulk of most eukaryote DNA and are classed as either Type I transposons (that have an RNA intermediate and can be thought of as 'copy and paste'), or Type II elements (where the segment of DNA can move, 'cut and paste'). piRNAs are processed from long endogenous predominantly antisense transposon or repeat sequence transcripts often found concentrated in hotspots on the genome.³⁹ Their biogenesis is different from the other pathways made possible because the Piwi clade of Argonautes have retained their catalytic capability. Although precursor piRNAs are mostly antisense, a small number of sense piRNAs are produced. These load into Ago3 and cleave antisense retrotransposon transcripts or the antisense precursor piRNA leaving a 5' antisense product or secondary piRNA. These load preferentially into Piwi or Aub and bind and cleave transposon mRNA forming a 5' sense piRNA that loads into Ago3 and guides cleavage of an antisense mRNA or the antisense precursor. This is known as a 'forward feed' or 'ping pong' amplification loop and can generate large numbers of piRNAs without the need for

Dicer processing. Plants and worms have also amplification steps in their siRNA pathways, generating secondary siRNAs though the mechanisms are quite different. Silencing is achieved principally by DNA methylation via the RITS (RNA-induced Initiation of Transcriptional Gene Silencing) complex containing Piwi and AUB. In mammals this epigenetic regulatory role is heritable through maternal transmission.⁴⁰ A comparison of some features of RNAi defence networks is shown in Figure 2.

Some genomic regions give rise to both endo-siRNA and piRNA and target the same transcripts, indicating cooperation/interaction between networks. Indeed, integration across the entire RNAi networks seems necessary, otherwise a viral infection might overwhelm the machinery required to suppress transposons to the mutual benefit of virus and transposon. Alternatively, there could be some redundancy among the molecular machinery that keeps all pathways working. For instance, it is known that the plant

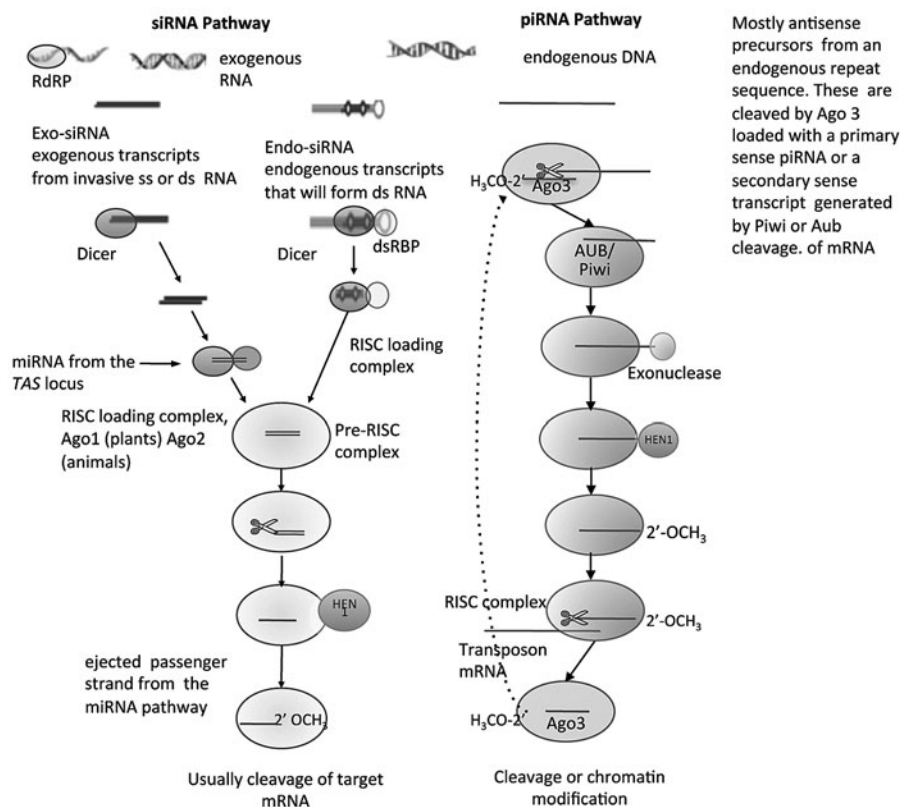


Figure 2. A comparison of RNAi networks involved with the defence of the Dark Arts. Current invasions can be quashed by exo-siRNAs utilising the RNA derived from the invader, either conveniently provided, for example by the geminivirus, or formed into double stranded RNA by RdRP. Endogenous invasive DNA is dealt with by the endo-siRNA pathway, sharing many proteins with the exo-siRNA and also the miRNA networks. The piRNA pathway is also adapted to prevent expression of endogenous invasive DNA, usually by modification of DNA architecture and specifically in the germline. Although the piRNA pathway is independent of Dicer it utilises a variant of the Argonaut protein confined to germline cells (see Table 1). A color version of this figure is available at www.landesbioscience.com/curie.

DCL2 can take over the functions of DCL4. Is this redundancy driven by fortuitous gene duplication, or in response to the plant dependence on the RNAi networks?

Viral particles evolve rapidly and even miRNA genes within related viruses may have little homology although they maintain sequence similarity with their mRNA targets for host mRNAs and with some cellular miRNAs. Viruses can suppress host miRNA and siRNA pathways via RNA silencing suppressors (RSS) at many steps. Some DNA viruses score 'own goals' by producing dsRNA, an example being the cassava mosaic virus, a geminivirus that produces dsRNA via convergent transcription from opposite promoters. The plant cytoplasmic DCL4 slices these into siRNAs that suppress the viral transcripts. However, the Dark Arts are never idle; viruses have evolved RSS to turn off the host defence, by inactivating DCL4. Nevertheless, the plant can retaliate and use DCL2 as a substitute for DCL4.⁴¹ The geminiviruses that give rise to convergent transcripts can inhibit the RdrP needed for the miRNA and siRNA pathways in plants,⁴² and prevent RNA-directed DNA methylation (Fig. 3). Plants have counter-evolved a method of methylating and silencing geminivirus minichromosomes using RNA-directed methylation via Ago4.⁴³ Geminivirus counteracts by suppressing methylation globally. A similar arms race occurs wherever viral nucleic acid is found, that is, everywhere. From

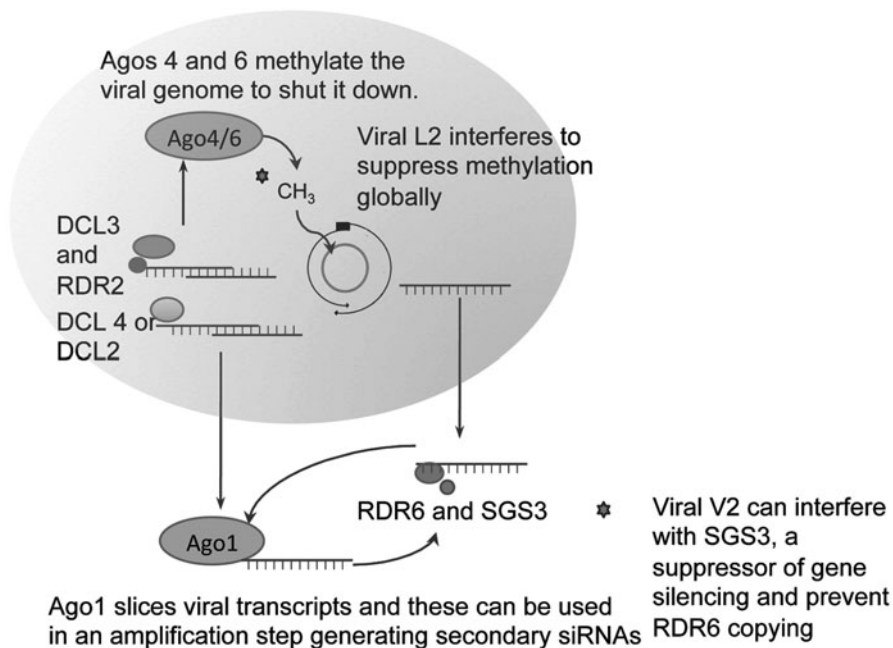


Figure 3. Geminiviruses are ssDNA viruses that replicate via rolling replication forming dsDNA without a dsRNA stage. Transcription is bidirectional and the transcript overlaps result in a dsRNA of a small segment of the *AC1* gene, which is indispensable for replication. DCL4 preferentially dices this into small (22nt) virus-derived RNAs (exo-siRNA) which load into Ago1 and slice mRNA transcripts arising from the *AC1* gene. The cleaved mRNA can be converted by plant RDR6 into secondary siRNAs in an amplification step, which also converts other viral ss transcripts into siRNAs. An alternate fate of the viral overlapping transcript is to be extended by RDR2 and diced by DCL3 in the nucleus, where it enters Agos 4 or 6 and shuts down viral replication. However, the viruses can interfere with this shutdown at different stages. A color version of this figure is available at www.landesbioscience.com/curie.

an evolutionary perspective the RNAi networks must be very old. Not only because RNAi is ubiquitous, but because it makes use of ancient proteins and is involved with the maintenance of other very old proteins such as the highly conserved histones which serve to enhance copying fidelity.

So what of bacteria and archaea, how does their arsenal stack up against parasitic nucleic acids? “Clustered, regularly interspaced, short palindromic repeats” (CRISPRs)⁴⁴ occur extensively in bacteria and archaea as a defence against phages, plasmids and transposons, using small RNAs to interfere with invasion. CRISPRs appear ubiquitous in archaea and the apparent lower incidence (~40%) in bacteria may be an artefact of lab maintained strains. However, an alternate explanation is that some pathogenic bacteria, reliant on plasmids for antibiotic resistance or virulence factors, have mechanisms to avoid the acquisition of CRISPR systems. Some *E. coli* strains have CRISPR arrays containing spacer sequences derived from *cas* genes termed ‘antiCRISPRs’ that can destroy the DNA of invading CRISPR/*cas* systems.⁴⁵ These species lack endogenous *cas* genes yet the antiCRISPR is selectively retained presumably to permit horizontal gene transfer, or for an as yet unidentified function. The CRISPR system is dynamic and evolutionary and an outline is given in Figure 4 (for more detail on the CRISPR system, please see Chapter 13, pages 213-218).

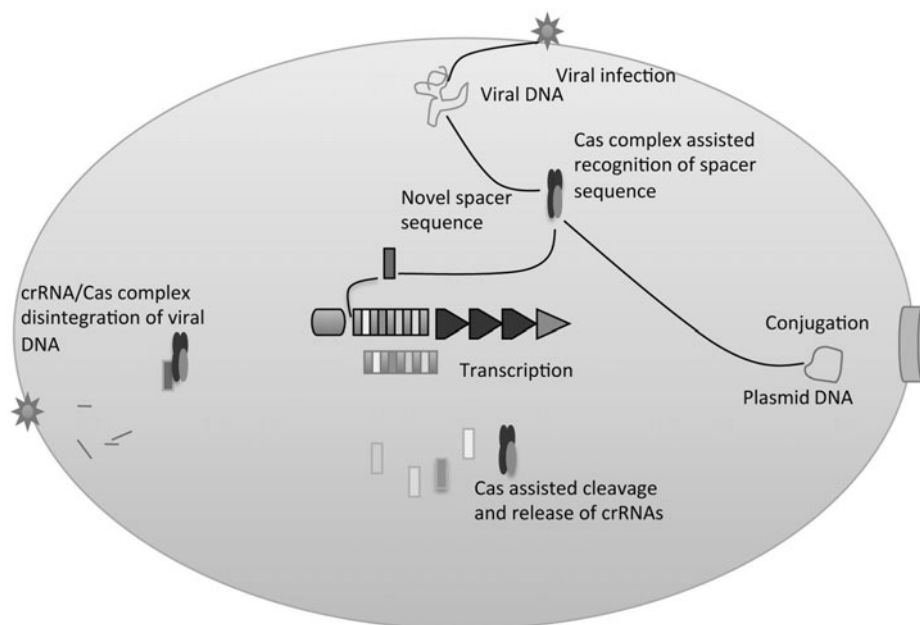


Figure 4. The CRISPR system is a combined exo and endo-siRNA that results in a dynamic, heritable immune system. crRNAs are produced from an endogenous transcript akin to endo-siRNA and diced using Cas proteins, these are analogous but not homologous to the Dicing proteins of the siRNA system. The spacer sequence is acquired from a current infection (top and right), though it isn't clear how rapidly this can be utilised, it would seem that it would be slower than eukaryote exo-siRNA due to the integration step. A major difference between the prokaryote and eukaryote defence against the Dark Arts is that evidence thus far would indicate that cleavage is directed against invading DNA (left), rather than mRNA. A color version of this figure is available at www.landesbioscience.com/curie.

Spacers between the CRISPR repeats are usually from sources external to the organism (e.g., from the phage) and possession of a spacer sequence matching a challenging phage gives resistance to that phage.⁴⁶ A suite of conserved proteins involved in DNA transactions was identified as typically occurring close to the CRISPR locus and called *cas* genes—(CRISPR associated). They have functional domains typical of nucleases, polymerases, helicases and nucleotide binding proteins⁴⁷ and are required for the acquisition of the spacer sequences. Although the mechanism is speculative, studies have shown that the core *cas* genes Cas1 and Cas2 occur at all CRISPR loci and are involved with acquisition of the spacer sequences. Disruption to these genes results in loss of ability to acquire new spacer sequences, but does not prevent the function of existing ones. Entire CRISPR/*cas* systems can also be found in plasmids,⁴⁸ and the same subset can be found in distantly related organisms. This indicates that they can be acquired from plasmids—by the very mechanism that they normally prevent.

Constitutive transcription of the entire repeat/spacer array gives long precursors which are processed into short crRNAs. Each crRNA corresponds to one spacer sequence flanked by two partial repeats.⁴⁷ The significance of the flanking sequences is that the resulting crRNAs can recognise ‘self’ DNA because the ends of the crRNAs will be able to attach by Watson-Crick pairing along their entire length, whereas predatory DNA will result in the crRNA being free at both ends.⁴⁹ Disruption of invading target DNA with a self-splicing intron that restores the mRNA, results in loss of resistance in *Staphylococcus epidermidis*, despite having the correct spacer sequence. This indicates that mRNA is not the target.⁵⁰ Bidirectional processing was found in the archaeon *Sulfolobus*, suggesting the possibility that RNA duplexes could form and given that archaea contain Argonautes, could this possibly represent a form of siRNA pathway akin to eukaryotes.

RNA cleavage has been demonstrated in vitro in another archaeon *Pyrococcus* via an RNP complex comprised of crRNA and Cas proteins (Cmr1-Cmr6) encoded from a Cas associated module known as RAMP (repeat associated mysterious protein, mainly found in thermophilic archaea).⁵¹ RAMP proteins can be located distally from the CRISPR and do not appear to move with the CRISPR array. DNA cleavage wasn't detected in *P. furiosus*, this may mean that the *Pyrococcus* CRISPR system is directed at mRNA, or directed at the genome of an RNA phage. The spacer sequences in *P. furiosus* have not been identified. If they are aimed at RNA phage, this is not surprising given that so few RNA phage have been sequenced and spacer sequences matching any RNA phage have not (yet) been found. It is possible that species with RAMP encoded Cas proteins have an alternate mechanism of silencing invasive nucleic acid.

For prokaryotes, with both a single center for DNA replication (Ori) and a short life span, it is vital that the genome is small so that replication can be carried out quickly.⁵² It would be costly to allow CRISPR arrays to grow unchecked and there is evidence of spacer deletion at the 3' end of the array.⁴⁷ This would allow historic invasions to be forgotten if the cost of maintaining the surveillance was high. There are also higher levels of expression of mature crRNA from the leader end of the transcript (from more recently acquired spacers) and lower levels of distal sequences.⁴⁷ Such a strategy ensures protection against current threats, whilst keeping a low level of surveillance for the re-emergence of an old foe. The Dark Arts cannot be so easily subdued. Mutations in a 3-5 proto-spacer adjacent motif (PAM) can prevent proto-spacer sequence recognition, or target sequence mutation can evade crRNA target recognition, phage from *Leptospirillum* can reshuffle 25 nt blocks⁵³ confounding the slightly larger spacer sequences of the crRNA. And so

it goes on. For the convenience of molecular evolutionists, CRISPRs give a historical insight into past host and pathogen encounters!

Thus it seems that all living cells can use some type of small RNA to defend themselves. Furthermore, we expect that there always would have been viruses, so the simplest hypothesis at present is that small RNAs have always been used in Defence Against the Dark Arts. Eukaryotes, bacteria and archaea have all evolved slightly different methods of dealing with parasitic DNA, but the use of RNA is a common thread. The diversity in mechanisms shows evolution has continued, but reinforces that the whole network system is very old.

OTHER REGULATORY RNAs

Our knowledge of noncoding RNA regulated pathways has expanded markedly in the past decade.⁵⁴⁻⁵⁶ The roles of networks of regulatory and catalytic ribonucleoprotein particles (RNPs) in eukaryotes have been well studied, including classical examples such as the small-nuclear RNPs (snRNPs) in mRNA splicing,^{57,58} small-nucleolar RNPs (snoRNPs) in rRNA processing—either the 2' hydroxyl of ribose is to be methylated, or a uracil converted to pseudouracil,⁵⁹ and the RNase P in tRNA processing.⁶⁰ Genome wide expression studies of eukaryotic model organisms have shown that over 90% of the genome is transcribed, which suggests that the richness of noncoding regulatory elements extends well beyond what is currently known.⁶¹

The conserved infrastructural network of RNP-mediated networks in eukaryotic cells indicates a pre-eukaryotic origin for many RNP functions.⁶² Both experimental and computational genome-wide mining of noncoding RNAs have revealed that many regulatory RNAs have highly flexible expression patterns.^{63,64} An example of such RNAs are snoRNAs that have been found in all eukaryotes and form one of the largest families of noncoding RNAs.³ They are not restricted to rRNA biogenesis, snoRNAs in the Cajal body (“small Cajal body RNAs”—scaRNAs)⁶⁵ modify snRNAs that function in the spliceosome. They have been identified as precursor for smaller RNAs;⁶⁶ a human HBII-52 snoRNA (SNORD 115) is processed into smaller RNAs, which regulate alternative splicing of the serotonin-receptor gene,⁶⁷ and alternative splicing itself appears very ancient in eukaryotes.⁶⁸ Similarly, a C/D box snoRNA from the Epstein-Barr virus is processed into smaller RNAs.⁶⁹ Many snoRNAs in modern genomes appear to arise by duplications.^{70,71} Large numbers of new snoRNAs are continuously being identified through experimental and bioinformatic screens.⁷² All these findings, together with the existence of many orphan snoRNAs without known targets⁷³ suggest additional network interactions. Identification of snoRNA-derived small RNAs from such a wide range of organisms also implies that some alternative regulatory roles of snoRNAs have an ancient origin. In some eukaryotic lineages with smaller genomes (such as diplomonads and microsporidia) and in archaea, there are fewer annotated snoRNAs, but those that are there appear functional.⁷⁴

Cis-regulatory RNAs have been identified in both eukaryotes and prokaryotes,³ and classic examples include riboswitches in bacteria and plants,⁷⁵ iron-response elements (IREs),⁷⁶ and the eukaryotic histone 3'UTR stem-loop.⁷⁷ Cis-regulatory RNAs are usually located in the untranslated regions (UTRs) in mRNAs, though some are in coding regions.⁷⁸ In association with RNA-binding proteins they regulate translation, splicing, stability and localization. For proteins, there are hundreds of identified RNA-binding protein domains,⁷⁹

but the identification of complementary protein-binding motifs in RNA is still at an early stage. An important role of cis-regulatory RNA elements is mRNA localization in eukaryotes. Expression of many eukaryotic RNAs involves packaging of mRNA-protein complexes into transport particles, trafficking of the particles within the cytoplasm and finally translated at their target destination. A few examples are localization of budding yeast *ASH1* mRNA to the bud tip,⁸⁰ localization of β -actin mRNA to the leading edge of migrating fibroblasts,⁸¹ localization of *Nonos* mRNA to the posterior end of the *Drosophila* embryo,⁸² and localization of human Vimentin mRNA to the perinuclear region of the cytoplasm.^{3,83} The noncoding motifs on the 3'-UTRs of the above mRNAs function to bind specific proteins for the transport of the mRNAs. Existence of these RNAs in all three kingdoms of life also hints at this network arising prior to the universal common ancestor (Luca).

HOW OLD ARE THE DIFFERENT INTERACTIONS OF RNA?

In general a good case can be made for many of the general classes of RNA networks to arise relatively early in evolution,^{84,85} but we need to outline the options carefully. Current approaches attribute importance to several overlapping earlier stages⁸⁶ during the origin of life (see Fig. 5). We start with the familiar world of DNA, proteins and RNA

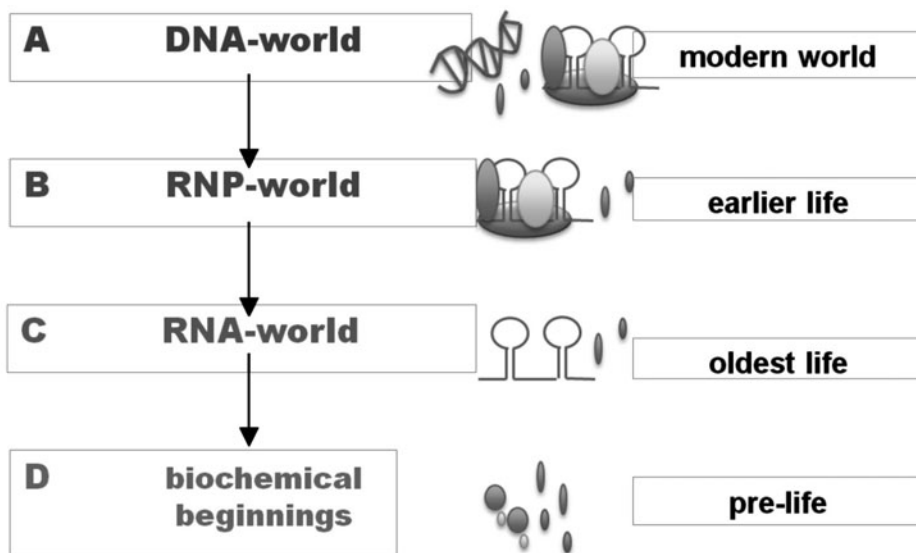


Figure 5. Working backwards through four stages for the origin of life. (A) Our modern world has DNA as the main information storage molecule and its double-stranded structure allows a higher replication accuracy. The standard model suggests an earlier RNA-protein (RNP) world (B) where RNA was the main molecule for genetic information and proteins had the main catalytic role. Many regulatory networks may date back to this time. At a still earlier stage (C), the standard model is an RNA-world where both catalysis and information storage was carried out by RNA—presumably aided by short (noncoded) peptides. The error rate is assumed to be relatively high and thus only relatively short RNAs could be coded. Earlier still there must have been a chemical stage (D) where perhaps autocatalytic cycles were important. It is not specified when membranes (allowing protocells) first arose. A color version of this figure is available at www.landesbioscience.com/curie.

and work backwards to earlier stages. The biological world that we are familiar with has DNA as the main coding molecule and this allows several steps of error checking (against the complementary strand as just one example). The replication accuracy of DNA-based system is much higher than for RNA-coding systems and as far as we can tell, the size of even the larger eukaryote genomes is no longer limited by the accuracy of replication of DNA.

Our modern world (Fig. 5A) has the three familiar groups of macromolecules—proteins, RNA and DNA. On any evolutionary scenario there will have been earlier and simpler worlds and under the standard theory it is predicted that our modern world was preceded (Fig. 5B) by a ribonuclear protein world (RNP) with RNA having the coding roles and proteins doing most of the catalysis. Part of the evidence that DNA evolves last comes from the necessity for protein enzymes (ribonucleotide reductases) that use free radical mechanisms to reduce an OH of ribose (in a ribonucleotide RNA precursor) to a deoxy-ribonucleotide.⁸⁷ As far as we know, only proteins (not RNA) can catalyse such complex free radical reactions. In the proposed ribonucleoprotein-world (RNP of Fig. 5B) protein is doing most catalysis (especially of small molecules), leaving RNA with the main coding function and RNA-protein networks would probably have had many regulatory roles.

From our knowledge of modern biochemistry, protein enzymes copying single-stranded RNA are much less accurate and have higher error rates than copying the double stranded DNA molecule. This lower accuracy means that RNA genomes would be much shorter than DNA-based organisms. Indeed, we see today that RNA viruses have much shorter genomes than double-stranded DNA viruses.⁸⁸ Effectively there is a positive feedback cycle of increasing fidelity of replication allowing longer coding sequences, which allows a further increase in replication fidelity—we call this the Darwin-Eigen cycle.⁸⁹ But the important point here is that from this stage we will have many RNA-protein interactions and so, in principle, some of the RNA-protein networks could have been established at this stage. Or were they lost in Luca and then later re-established—but only in eukaryotes? Losing them and later re-establishing them, seems a less likely hypothesis.

At this point (ribonucleic acids and proteins) we are still on reasonably firm ground; but what comes earlier? Currently the best hypothesis for the earlier stage (Fig. 4C) is that RNA is the first polymer that had the ability to reproduce itself (either directly, or indirectly through a cycle and with or without assistance of short noncoded peptides). It has long been considered⁹⁰ that enzyme cofactors with a ribodinucleotide structure (such as NAD, FAD, NADP) were remnants of an early RNA-world. The evidence for this hypothesis of the early role of RNA is increasing.⁹¹ In this proposed RNA-world, RNA would have had coding, catalytic and regulatory roles. An interesting point is that some RNAs and their interactions, appear to date from this proposed RNA-world. RNA is not as an effective catalyst as proteins⁸⁵ (probably because proteins form more specific and stable 3D structures) and so this leaves space for improved protein catalysts in going from the second to the third stage of Figure 5. There had to be earlier stages before organised polymers formed (whether RNA, protein, or DNA) and we summarize these stages as 'biochemical beginnings' (Fig. 5D).¹ Basically by definition, we do not expect any macromolecules to have persisted from this stage.

Some classes of functional and regulatory RNA are so widespread that they are accepted as 'universal' and are inferred to have been present in the last common ancestor, Luca. These include rRNA, tRNA, mRNA, RNase P, SRP RNA. A striking discovery was that the catalytic core of the ribosome (forming the amino-acyl bonds linking amino acids into

proteins) was composed of RNA. In other words, the ribosome was a ribozyme.⁹² Thus the three core aspects of protein synthesis—the messenger, transfer RNA and ribosome—are all RNA molecules that have a network of interactions. Evolutionarily this makes sense if an RNA-world preceded an RNP-world. However there are few convincing theories for the origin of protein synthesis because we need these three classes of RNA before proteins can be synthesized. In an evolutionary context, we cannot evolve something ‘because it will be useful in the distant future’—it must have a function ‘here and now’. Certainly rRNA, tRNA and mRNA must date back to an RNA-world (before proteins). Sometimes we can define useful subquestions—such as whether the intron/exon structure, with its associated RNAs involved in splicing, is ancestral to all modern eukaryotes.⁹³ Some RNAs and their interactions may yield progress relatively easily; others will be more difficult for now.

A recent discovery is that tRNAs are part of the proof-reading that increases accuracy,⁹⁴ the tRNA(Leu) has catalytic activity in removing a mischarged tRNA(Leu), even though protein does help stabilise some of the intermediates along the catalytic pathway. This means that we need to consider that tRNAs are not passive in amino acid recognition; they are also part of the regulatory network. Although not so well known as the RNAs involved in protein synthesis, the signal recognition particle, 7S in eukaryotes,⁹⁵ also appears to occur in all living systems, so this RNA-protein network also appears ancient.

Were these RNA interactions all present in Luca? We will not follow this question here, but have written on it⁸⁹ under the names introns ‘first’, ‘early’ or ‘late’. Because the position of the root of the eukaryote tree is not certain,⁹⁶ our strategy⁹³ has been to search for RNAs that occur in all of the five (or six) deepest lineages of eukaryotes. For example, the intron/exon structure of eukaryote genomes, with its requirement for splicing, appears universal in all modern eukaryotes, so it is expected to have been in the last common eukaryote ancestor (Fred). But that only puts the problem back further. For example, did snRNAs arise before, or after, another milestone—the endosymbiotic event that led to the origin of mitochondria? Details on the relative ages of classes will depend on the models of genome reduction.⁹⁷ There are plenty of interesting and fundamental questions for the next decade.

CONCLUSION

Every eukaryote group appears to have networks that handle small RNAs a little differently. However, this is likely a normal evolutionary process creating variants arising from an early ancestral system in Fred. Perhaps it is still premature to decide how the ancestral eukaryote functioned in this respect, but the discovery of different forms of RNA networks in so many groups of eukaryotes has turned the focus back to earlier stages.

Parasitic RNA would have been present in the RNA and RNP-worlds; that is the nature of biology. Thus we expect that at least from the time of Luca there would have been mechanisms to protect the early genome from parasites of various kinds—whether viral or transposon-like. We expect that the Defence Against the Dark Arts idea would have always been a problem for cells since they all require some type of defence mechanism. Although the machinery differs, prokaryotes and eukaryotes have networks with roughly the same function using small RNAs to keep the Dark Arts at bay. It seems likely that the siRNA system was present in Luca as the Argonautes are in all three domains and

are almost universally conserved, indicating their age and importance. The endo and piRNA pathways may have evolved to suppress parasitic nucleic acid that escaped or overrun the exo-siRNA pathway.

Increasingly, more RNAs seem to be universal. Bacteria and viruses have many cis-regulatory elements which bind proteins or metabolites (riboswitches), plants and fungi have riboswitches and now it looks like that cis-regulatory elements (in 5'- and 3'-UTRs) are very common in eukaryotes too. Large scale transcription studies in human and mouse at least have uncovered many UTR-derived RNAs, many of which are likely to contain cis-regulatory elements and they may work in a similar way as those in bacteria and virus (binding to specific proteins and having regulatory roles in transcription).

At present we must keep an open mind about how old the RNA and proteins systems are that form the RNA networks. At present, we favour the simplest hypothesis that the RNA networks are, in their basic form, very old. But we must keep testing this hypothesis, looking for small RNA networks involved in regulation in as wide a range of groups as possible. It is a stimulating time for research on RNA networks.

ACKNOWLEDGEMENTS

We thank Lesley Collins for important editing and comments for this chapter.

REFERENCES

1. Yarus M. Getting past the RNA world: The initial Darwinian ancestor. Cold Spring Harbor Perspectives in Biology 2010. doi: 10.1101/cshperspect.a003590.
2. Penny D, Collins LJ. Evolutionary genomics leads the way. In: Caetano-Anolles, ed. Evolutionary Genomics and Systems Biology. Hoboken: Wiley-Blackwell, 2010:3-16.
3. Gardner PP, Daub J, Tate JG et al. Rfam: updates to the RNA families database. Nucl Acids Res 2009; 37:D136-D140.
4. Boria I, Gruber AR, Tanzer A et al. Nematode sbrRNAs: homologs of vertebrate Y RNAs. J Mol Evol 2010; 70:346-358.
5. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Genet 2009; 10:102-108.
6. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev 2004; 5:396-400.
7. Carlsbecker A, Lee JY, Roberts CJ et al. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature 2010; 465:316-321.
8. Cullen BR. Viral and cellular messenger RNA targets of viral microRNAs. Nature 2009; 457:421-425.
9. Ouellet DL, Plante I, Landry P et al. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucl Acids Res 2008; 36:2353-2365.
10. Lin J, Cullen BR. Analysis of the interaction of primate retroviruses with the human RNA interference machinery. J Virol 2007; 81:12218-12226.
11. Dunoyer P, Himber C, Voinnet O. Induction, suppression and requirement of RNA silencing pathways in virulent agrobacterium tumefaciens infections. Nat Genet 2006; 38:258-263.
12. Dong Z, Han MS, Federoff N. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. PNAS 2008; 105:9970-9975.
13. Okamura K, Liu N, Lai EC. Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. Mol Cell 2009; 36:431-444.
14. Berezikov E, Chung W, Willis J et al. Mammalian mirtron genes. Mol Cell 2007; 28:328-333.
15. Zamore PD, Haley B. Ribo-gnome: the big world of small RNAs. Science 2005; 309:1519-1524.
16. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. Nature 2009; 457:413-420.
17. Carthew RW, Sonheimer EJ. Origins and mechanisms of miRNAs and siRNAs. Cell 2009; 136:642-655.
18. Leung AKL, Sharp PA. MicroRNAs: A safeguard against turmoil? Cell 2007; 130:581-585.

19. Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008; 30:460-471.
20. Lee YS, Shibata Y, Malhotra A et al. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes and Dev* 2009; 23:2639-2649.
21. Haussecker D, Huang Y, Lau A et al. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 2010; 16:673-695.
22. Kulkarni M, Ozgur S, Stoecklin G. On track with P-bodies. *Biochem Soc Trans* 2010; 38:242.
23. Guang S, Bochner AF, Pavelec DM et al. An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* 2008; 321:537-541.
24. Bartel B. MicroRNAs directing siRNA biogenesis. *Nat Struct Mol Biol* 2005; 12:569-571.
25. Miyoshi K, Miyoshi T, Hartig JV et al. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* 2010; 16:506-515.
26. Robine N, Lau NC, Balla S et al. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* 2009; 19:2066-2076.
27. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell* 2009; 136:656-668.
28. Li C, Vagin VV, Lee S et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 2009; 137:509-521.
29. Gottesman S. Micros for microbes: noncoding regulatory RNAs in bacteria. *Trends Genet.* 2005; 21:399-404.
30. de Noijer S, Holland BR, Penny D. Eukaryote origins: there was no Garden of Eden? *PLoS One* 2009; 4:e5507.
31. Boerlijst MC, Hogeweg P. Spiral wave structure in prebiotic evolution—hypercycles stable against parasites. *Physica D* 1991; 48:17-28.
32. Yigit E, Batista PJ, Bei Y et al. Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 2006; 127:747-757.
33. de Vries W, Berkhout B. RNAi suppressors encoded by pathogenic human viruses. *Int J Biochem Cell Biol* 2008; 40:2007-2012.
34. Parameswaran P, Sklan E, Wilkins C et al. Six RNA viruses and forty-one hosts: viral small RNAs and modulation of small RNA repertoires in vertebrate and invertebrate systems. *PLoS Pathogens* 2010; 6:e1000764.
35. Ruiz-Ferrer V, Voinnet O. Roles of plant small RNAs in biotic stress responses. *Annu Rev Plant Biol* 2009; 60:485-510.
36. Lipardi C, Paterson BM. Identification of an RNA-dependent RNA polymerase in *Drosophila* involved in RNAi and transposon suppression. *PNAS* 2009; 106:15645-15650.
37. Maida Y, Yasukawa M, Furuuchi M et al. An RNA dependent RNA polymerase formed by hTERT and the RNase MRP RNA. *Nature* 2009; 461:230-235.
38. Watanabe T, Totoki Y, Toyoda A et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 2008; 453:539-543.
39. Hartig JV, Tomari Y, Forstemann K. piRNAs—the ancient hunters of genome invaders. *Genes Dev* 2007; 21:1707-1713.
40. Brennecke J, Malone CD, Aravin AA et al. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 2008; 322:1387-1392.
41. Ding S, Voinnet O. Antiviral immunity directed by small RNAs. *Cell* 2007; 130:413-426.
42. Wang H, Buckley KJ, Yang X et al. Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. *J Virol* 2005; 79:7410-7418.
43. Raja P, Sanville BC, Buchmann RC et al. Viral genome methylation as an epigenetic defence against geminiviruses. *J Virol* 2008; 82:8997-9007.
44. Jansen R, Embsen JD, Gastra W et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002; 43:1565-1575.
45. Touchan M, Rocha EPC. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PloS One* 2010; 5(6):e11126.
46. Bolotin A, Quinquis B, Sorokin A et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005; 151:2551-2561.
47. Wiedenheft B, Zhou K, Jinek M et al. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 2009; 17:904-912.
48. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 2008; 320:1047-1050.
49. Marraffini LA, Sontheimer EJ. Self versus nonself discrimination during CRISPR RNA-directed immunity. *Nature* 2010; 463:568-571.
50. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 2008; 322:1843-1845.

51. Hale C, Kleppe K, Terns RM et al. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 2008; 14:2572-2579.
52. Penny D, Poole AM. Lateral gene transfer: some theoretical aspects. *NZ BioScience* 2003; 12:32-35.
53. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 2008; 320:1047-1050.
54. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* 2010; 11:75-87.
55. Lioliou E, Romilly C, Romby P et al. RNA-mediated regulation in bacteria: from natural to artificial systems. *New Biotech* 2010; 27:222-235.
56. Sashital DG, Doudna JA. Structural insights into RNA interference. *Curr Opin Struct Biol* 2010; 20:90-97.
57. Newman AJ, Nagai K. Structural studies of the spliceosome: blind men and an elephant. *Curr Opin Struct Biol* 2010; 20:82-89.
58. Staley JP, Woolford JL Jr. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr Opin Cell Biol* 2009; 21:109-118.
59. Bachelier JP, Cavaille J, Huttenhofer A. The expanding snoRNA world. *Biochimie* 2002; 84:775-790.
60. Lai LB, Vioque A, Kirsebom et al. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett* 2010; 584:287-296.
61. Carninci P. RNA dust: where are the genes? *DNA Res* 2010; 17:51-59.
62. Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet* 2009; 25:120-128.
63. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 2009; 94:83-88.
64. Slezak-Prochazka I, Durmus S, Kroesen BJ et al. MicroRNAs, macrocontrol: regulation of miRNA processing. *RNA* 2010; 16:1087-1095.
65. Kiss AM, Jady BE, Darzacq X et al. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucl Acids Res* 2002; 30:4643-4649.
66. Taft RJ, Glazov EA, Lassmann T et al. Small RNAs derived from snoRNAs. *RNA* 2009; 15:1233-1240.
67. Kishore S, Khanna A, Zhang et al. The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet* 2010; 19:1153-1164.
68. Irimia M, Rukov JL, Penny D et al. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol Biol* 2007; 7:188.
69. Hutzinger R, Feederle R, Mrazek J et al. Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome. *PLoS Pathog* 2009; 5:e1000547.
70. Chen CL, Liang D, Zhou et al. The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucl Acids Res* 2003; 31:2601-2613.
71. Zemmann A, op de Bekke A, Kiefmann M et al. Evolution of small nucleolar RNAs in nematodes. *Nucl Acids Res* 2006; 34:2676-2685.
72. Raabe CA, Sanchez CP, Randau et al. A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. *Nucl Acids Res* 2010; 38:608-617.
73. Hertel J, Hofacker IL, Stadler PF. SnoReport: computational identification of snoRNAs with unknown targets. *Bioinformatics* 2008; 24:158-164.
74. Gardner PP, Bateman A, Poole AM. SnoPatrol: how many snoRNA genes are there? *J Biol* 2010; 9:4.
75. Henkin TM. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* 2008; 22:3383-3390.
76. Pantopoulos K. Iron metabolism and the IRE/IRP regulatory system: an update. *Ann N Y Acad Sci* 2004; 1012:1-13.
77. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* 2007; 396:373-390.
78. Nguyen MQ, Zhou Z, Marks CA et al. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* 2007; 131:1009-1017.
79. Finn RD, Mistry J, Tate J et al. The Pfam protein families database. *Nucl Acids Res* 2010; 38:D211-D222.
80. Gonsalvez GB, Urbinati CR, Long RM. RNA localization in yeast: moving towards a mechanism. *Biol Cell* 2005; 97:75-86.
81. Condeelis J, Singer RH. How and why does beta-actin mRNA target? *Biol Cell* 2005; 97:97-110.
82. Kugler JM, Lasko P. Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during *Drosophila* oogenesis. *Fly* 2009; 3:15-28.
83. Bermano G, Shepherd RK, Zehner ZE et al. Perinuclear mRNA localisation by vimentin 3'-untranslated region requires a 100 nucleotide sequence and intermediate filaments. *FEBS Lett* 2001; 497:77-81.
84. Collins LJ, Chen XS. Ancestral RNA: the RNA biology of the eukaryotic ancestor. *RNA Biol* 2009; 6:1-8.
85. Jeffares DC, Poole AM, Penny D. Relics from the RNA world. *J Mol Evol* 1998; 46:18-36.
86. Penny D. An interpretive review of the origin of life research. *Biol Philos* 2005; 20:633-671.

87. Poole A, Penny D, Sjöberg BM. Confounded Cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2001; 12:147-151.
88. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 2008; 9:267-276.
89. Penny D, Hoepfner MP, Poole AM et al. An overview of the introns-first theory. *J Mol Evol* 2009; 69:527-540.
90. White HB. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* 1976; 7:101-104.
91. Yarus M. *Life from an RNA World: The Ancestor Within*. Cambridge:Harvard University Press, 2010.
92. Steitz TA, Moore PB. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* 2003; 28:411-418.
93. Collins LJ, Penny D. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol* 2005; 22:1053-1066.
94. Hagiwara Y, Field MJ, Nureki O et al. Editing mechanism of aminoacyl-tRNA synthetases operates by a hybrid ribozyme/protein catalyst. *J Am Chem Soc* 2010; 132:2751-2758.
95. Rosenblad MA, Gorodkin J, Knudsen B et al. SRPDB: signal recognition particle database. *Nucl Acids Res* 2003; 31:363-364.
96. Keeling PJ, Burger G et al. The tree of eukaryotes. *Trends Ecol Evol* 2005; 20:670-676.
97. Poole AM, Penny D. Lateral gene transfer, some theoretical aspects. *NZ BioScience* 2003:32-35.